

1 **CLAIMS**

2 1. A method comprising:

3 capturing a web page from a web site;

4 extracting data from the web page using a data harvesting script;

5 normalizing the extracted data; and

6 storing the normalized data in a database.

7
8 2. A method as recited in claim 1 wherein the web site is associated
9 with a financial institution.

10
11 3. A method as recited in claim 1 wherein the captured web page
12 contains information regarding a customer's account at the financial institution.

13
14 4. A method as recited in claim 1 wherein the web page is an HTML
15 screen.

16
17 5. A method as recited in claim 1 further comprising:

18 capturing a second web page from a second web site;

19 extracting data from the second web page using the data harvesting script;

20 normalizing the data extracted from the second web page; and

21 storing the normalized data from the second web page in the database.
22
23
24
25

1 6. A method as recited in claim 1 further comprising generating an error
2 message if data cannot be extracted from the web page using the data harvesting
3 script.

4
5 7. A method as recited in claim 1 further comprising storing a copy of
6 the captured web page if data cannot be extracted from the web page using the
7 data harvesting script.

8
9 8. One or more computer-readable memories containing a computer
10 program that is executable by a processor to perform the method recited in claim
11 1.

12
13 9. A method comprising:
14 retrieving financial data associated with a user's financial account from a
15 data source;
16 identifying data of interest retrieved from the data source;
17 normalizing the identified data; and
18 storing the normalized data in a database.

19
20 10. A method as recited in claim 9 further comprising:
21 retrieving financial data associated with a user from a second data source;
22 normalizing the data retrieved from the second data source; and
23 storing the normalized data in the database.

1 **11.** One or more computer-readable memories containing a computer
2 program that is executable by a processor to perform the method recited in claim
3 9.

4
5 **12.** A method comprising:
6 capturing a web page from a web site;
7 attempting to extract data from the web page using a data harvesting script;
8 removing personal information from the captured web page;
9 storing the captured web page without the personal information; and
10 if data cannot be extracted from the web page, analyzing the web page and
11 the data harvesting script to determine why data could not be extracted from the
12 web page.

13
14 **13.** A method as recited in claim 12 further comprising editing the data
15 harvesting script based on an analysis of the captured web page.

16
17 **14.** A method as recited in claim 12 further comprising:
18 editing the data harvesting script based on an analysis of the captured web
19 page;
20 capturing a new version of the web page from the web site; and
21 extracting data from the web page using the edited data harvesting script.

22
23 **15.** A method as recited in claim 14 further comprising:
24 normalizing the data extracted from the web page; and
25 storing the normalized data in a database.

1
2 **16.** One or more computer-readable memories containing a computer
3 program that is executable by a processor to perform the method recited in claim
4 12.

5
6 **17.** A method comprising:
7 capturing a first web page from a first financial institution web site;
8 capturing a second web page from a second financial institution web site;
9 extracting data from the first web page using a first data harvesting script;
10 extracting data from the second web page using a second data harvesting
11 script;
12 normalizing the data extracted from the first web page and the second web
13 page; and
14 storing the normalized data in a database.

15
16 **18.** A method as recited in claim 17 further comprising generating an
17 error message if data cannot be extracted from the first web page or the second
18 web page.

19
20 **19.** One or more computer-readable memories containing a computer
21 program that is executable by a processor to perform the method recited in claim
22 17.

1 **20.** A method as recited in claim 17 wherein capturing a first web page
2 includes capturing a first set of web pages.

3
4 **21.** A method as recited in claim 17 wherein capturing a second web
5 page includes capturing a second set of web pages.

6
7 **22.** An apparatus comprising:
8 a data capture module configured to capture a web page from a web site
9 associated with a financial institution;
10 a data extraction module coupled to the data capture module and configured
11 to extract data from the captured web page using a data harvesting script, the data
12 extraction module further configured to normalize the extracted data; and
13 a database control module coupled to the data extraction module and
14 configured to store the normalized data in a database.

15
16 **23.** An apparatus as recited in claim 22 wherein the data capture module
17 is further configured to retrieve financial data associated with a user's account
18 from a data source.

19
20 **24.** An apparatus as recited in claim 22 wherein the data extraction
21 module is further configured to generate an error message if data cannot be
22 extracted from the web page using the data harvesting script.

1
2 **25.** One or more computer readable media having stored thereon a
3 plurality of instructions that, when executed by a processor, causes the processor
4 to perform acts comprising:

5 capturing a web page from a financial institution web site;
6 attempting to extract data from the captured web page using a data
7 harvesting script;
8 removing personal information from the captured web page;
9 storing the captured web page without the personal information; and
10 if data cannot be extracted from the web page, analyzing the web page to
11 determine why data could not be extracted from the web page.
12

13 **26.** One or more computer readable media as recited in claim 25,
14 wherein if data cannot be extracted from the web page, editing the data harvesting
15 script based on an analysis of the captured web page.
16

17 **27.** One or more computer readable media as recited in claim 25, further
18 comprising:

19 normalizing the data extracted from the web page; and
20 storing the normalized data in a database, wherein the database contains
21 data extracted from other web pages.
22
23
24
25